Robust signal extraction for on-line monitoring data

P.L. Davies[a], R. Fried[b], U. Gather[b,*]

[a]*Fachbereich Mathematik, Universität Essen, D-45141 Essen, Germany*

[b]*Fachbereich Statistik, Universität Dortmund, D-44221 Dortmund, Germany*

Abstract

Data from the automatic monitoring of intensive care patients exhibits trends, outliers, and level changes as well as periods of relative constancy. All this is overlaid with a high level of noise and there are dependencies between the different items measured. Current monitoring systems tend to deliver too many false warnings which reduces their acceptability by medical staff. The challenge is to develop a method which allows a fast and reliable denoising of the data and which can separate artifacts from clinical relevant structural changes in the patients condition (Gather et al., 2002). A simple median filter works well as long as there is no substantial trend in the data but improvements may be possible by approximating the data by a local linear trend. As a first step in this programme the paper examines the relative merits of the L_1 regression, the repeated median (Siegel, 1982) and the least median of squares (Hampel, 1975, Rousseeuw, 1984). The question of dependency between different items is a topic for future research.

MSC: primary: 62G07; secondary: 62G35

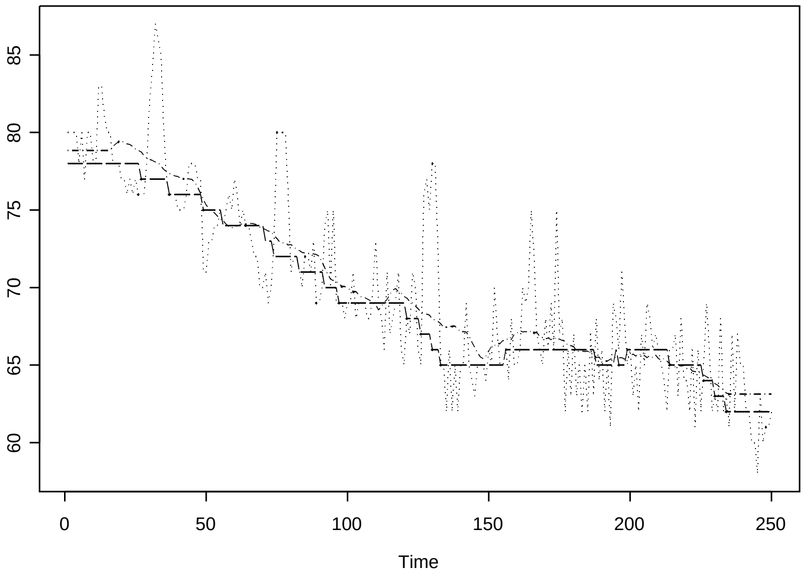
Keywords: Linear regression; Signal extraction; Level change; Trend; Outliers; Small-sample efficiency

1 Introduction

On-line monitoring of intensive care patients poses an interesting challenge for statisticians. Figure 1 shows a small excerpt from a series of measurements of the heart rate of a critically ill patient. An experienced physician analysed the data as being composed of a downward trend until time point 120 with noise and many clinically irrelevant outliers. Figure 1 also shows the result of a running mean and of a running median (Tukey, 1977) with a time window of 31 observations. Although both methods provide denoising, the mean is clearly effected by the outliers. The median resists the outliers much more successfully but approximates the more or less linear trend by a step function.

The superiority of the median in resisting the clinically irrelevant outliers indicates the advantages of robust statistical functionals. It seems plausible that the difficulties of the median in adapting to local trends can be overcome by the use of robust regression functionals. As a first step in this programme we investigate the relative

Figure 1: Time series of the heart rate (dotted), as well as a running mean (dashed) and a running median (solid) with window width 31 both.



merits of the L_1 regression, the repeated median and the least median of squares. Of particular interest are

- their ability to reproduce a linear trend in the presence of outliers
- their ability to detect trend changes
- their ability to detect level changes
- the cost of computation.

Traditionally the question of efficiency is also considered and we include some simulations for completeness. The important properties are however those listed above and these have little to do with efficiency (Davies and Gather, 1993). The situation we consider is a special one. The design points form a lattice and the sample size of about 20 to 30 observations is rather small but is necessitated by the requirement of being on-line. Clearly the more time one has the better the retrospective results but then it might be too late for the patient.

2 Methods for robust linear regression

The robustification of even the simple linear regression model $y = a + bx + \epsilon$ poses a considerable problem. One main weakness of all known high breakdown methods is their computational complexity. Huber (1995) has expressed this rather pointedly by saying that the high breakdown methods themselves break down because of their incomputability. The Hampel-Rousseeuw LMS functional T_{LMS} (Hampel, 1975, Rousseeuw, 1984) is defined by

$$T_{LMS} = \text{argmin}\{(a, b) : \text{Median}(y_i - a - bx_i)^2\}. \tag{1}$$

As the design points lie on a lattice a breakdown can only be caused by outliers in the y variable. In this situation the breakdown point of T_{LMS} in case of a sample of size n is $\lfloor n/2 \rfloor / n$. The computational complexity of T_{LMS} is of order n^4 (Stromberg, 1993). This can be reduced but only at the cost of attaining some approximation to the correct solution. As such approximations are unlikely to be permitted in on-line monitoring we are obliged to calculate the exact solution. As we are dealing with sample sizes of the order of 20 or 30 this complexity is no great problem for a single time series. If however several hundred items have to be treated simultaneously then the computational complexity may become a problem. In principle (1) may not have a unique solution but this has not proved to be a problem in practice.

Another high breakdown regression functional is Siegel's repeated median T_{RM} defined by

$$\begin{aligned} \tilde{\beta}_{RM} &= med_i \left(med_{j \neq i} \frac{y_i - y_j}{x_i - x_j} \right), \\ \tilde{\mu}_{RM} &= med_i (y_i - \tilde{\beta}_{RM} x_i) \, . \end{aligned}$$

Its breakdown point is also $\lfloor n/2 \rfloor / n$ and the computational complexity of it is of order n^2. It may therefore be preferred to T_{LMS} even if its small sample performance should turn out to be worse.

Finally we also consider the L_1 regression T_{L1} defined by

$$T_{L1} = \text{argmin}\{(a, b) : \sum_{i=1}^{n} |y_i - a - bx_i|\} \, . \tag{2}$$

The L_1 regression functional is the one most susceptible to outliers of the three methods. We calculate T_{L1} using the descent technique as described in Sposito (1990) which is slightly faster than other methods for sample sizes $n \leq 30$ and considerably faster for larger sample sizes. The existence of multiple solutions is a noticeable problem of T_{L1} but we shall ignore it. The breakdown point of T_{L1} can be calculated from the results of He et al.(1990) and Mizera and Müller (1999). For design points on a lattice, $x_1 = -m, \ldots, x_n = m$, i.e. $n = 2m + 1$, it reduces to

$$\min\left\{\frac{|I|}{2m+1} : \sum_{t\in I}|t| \geq \sum_{t\in I^c}|t|, I \subset \{-m,\ldots,m\}\right\}.$$

For large m, this is approximately $1 - 1/\sqrt{2} \approx 0.293$. Table 1 gives the exact values for small m.

Clearly T_{L1} is much less robust than either T_{LMS} or T_{RM} but its speed of calculation may make it an interesting candidate if several hundred regressions have to be performed simultaneously. For this reason we include it in the comparison.

Table 1: Finite-sample replacement breakdown point q_m of T_{L1}

m	1	2	3	4	5	6	7	8	9	10	11	12
q_m	1/3	2/5	2/7	3/9	4/11	4/13	5/15	5/17	6/19	7/21	7/23	7/25
$\approx$	0.33	0.40	0.28	0.33	0.36	0.31	0.33	0.29	0.32	0.33	0.30	0.28

3 Comparison of T_{L1}, T_{RM} and T_{LMS}

3.1 The basic simulation model

For a comparison of the finite-sample properties of the distinct regression methods 10000 samples were simulated using the model

$$Y_t = \mu + \beta t + \epsilon_t, \quad t = -m, \ldots, m,$$

with $\mu = 0$ and for several different slopes β. The error ϵ was always Gaussian white noise with mean zero and unit variance. The estimated values of μ and β are used to provide a value of the signal at time $t = 0$. This represents a time delay of m. The value of m is determined by requiring on the one hand a certain stability (m large) and on the other hand the demands made by the on-line nature of the application (m small). In this paper we restrict attention to the cases $m = 5, 10, 15$ which correspond to sample sizes $n = 2m + 1 = 11, 21, 31$.

3.2 Efficiency

As a first step we give the relative efficiencies of the three functionals T_{L1}, T_{RM} and T_{LMS} with respect to the least squares functional T_{L2}. As mentioned above efficiency

Table 2: Efficiencies relative to L_2 regression (in percent) measured by the simulated MSE for T_{L1} $(\tilde{\mu}_{L1}, \tilde{\beta}_{L1})$, T_{RM} $(\tilde{\mu}_{RM}, \tilde{\beta}_{RM})$ and T_{LMS} $(\tilde{\mu}_{LM}, \tilde{\beta}_{LM})$ for $N(0,1)$ errors.

m	β	$\tilde{\mu}_{L1}$	$\tilde{\mu}_{RM}$	$\tilde{\mu}_{LM}$	$\tilde{\beta}_{L1}$	$\tilde{\beta}_{RM}$	$\tilde{\beta}_{LM}$
5	0.0	69.7	66.3	26.9	78.8	69.8	25.1
5	0.1	71.1	66.4	27.4	71.9	70.4	26.1
5	0.2	69.3	64.3	26.3	63.6	68.8	24.7
10	0.0	66.9	63.9	22.4	70.4	70.8	22.7
10	0.1	67.8	64.4	22.9	64.5	71.7	23.4
10	0.2	69.4	66.6	23.1	66.1	73.1	24.2
15	0.0	66.3	64.3	20.7	70.2	71.4	21.6
15	0.1	68.1	65.0	20.7	64.5	72.7	21.0
15	0.2	68.0	65.4	20.4	66.1	73.2	22.0

is not an overriding consideration here. The results are given in Table 2 for the slopes $\beta = 0, \beta = 0.1, \beta = 0.2$. They show no great surprise except for the behaviour of the slope component of T_{L1} where the relative efficiency is highest for $\beta = 0$. This may well be due to the non-uniqueness of the L_1 solution and the result of taking $\beta = 0$ as a starting point for the calculation of the solution. A similar phenomenon was noted by Terbeck (1996) in the case of the two-way-table.

3.3 Outliers in the steady state

Data in intensive care medicine contain large isolated outliers as well as patches of outliers. For the sake of brevity we restrict attention to a sample size $n = 21$ and replace an increasing number of observations 0(1)10 by additive outliers of increasing size 0(2)10 at random points in the window. We concentrate on one-sided positive outliers as those constitute a difficult challenge and are more common than negative ones in intensive care. The simulations were performed with $\mu = \beta = 0$. Each of the 121 cases is simulated 2000 times and the squared bias, variance and mean square error were calculated. The results are shown graphically in Figure 2 for the intercept and in Figure 3 for the slope. For the latter, only the MSE is shown as outliers occurring at positions chosen at random do not cause a bias for the slope. For 0-6 outliers or for outliers of size 0-4 there is little to choose for the methods. T_{L1} shows considerable bias in the intercept for 7 or more outliers. This corresponds well with Table 1. T_{RM} performs similarly like T_{L1} for the intercept although it has the same breakdown point as T_{LMS}. Both T_{RM} and T_{L1} are dominated by T_{LMS} in the intercept for 8 or more outliers of any size. With respect to the slope, T_{RM} has the smallest MSE among the three functionals in case of many small outliers, while T_{LMS} is better for many large outliers.

3.4 Level shift and outliers

A situation that is particularly important in on-line monitoring is the occurrence of a level shift. In order to detect a level shift we need a reliable approximation of the current level when the last observations in the time window are at another level. Clearly some definition of a level shift is required to distinguish it from a block of outliers. The definition we take is that the last five observations are of about the same size and differ substantially from the preceeding observations (cf. Imhoff et al., 1998, Gather et al., 2000). We define "substantially" to be a difference in level of size $\omega \in \{3.0, 5.0, 10.0\}$. To provide a greater challenge we also put some positive outliers at time points chosen at random. Again the squared bias, variance and mean square error were calculated for each of the three regression functionals. The results for the shift $\omega = 10$ are shown in Figures 4 (intercept) and 5 (slope). Five outliers occurring at the end of the time window cause T_{L1} to be biased for the intercept and the slope, while T_{RM} is biased for the slope. The superiority of T_{LMS} is apparent. It shows much less bias than the other functionals and can even accommodate up to 7 outliers. The slope component of T_{LMS} shows considerable variability if a level shift and eight or more outliers occur. The results are of course less clear cut for a smaller positive shift but in this case T_{LMS} is again superior. In the case of the slope a moderate number of positive outliers and a negative shift can balance each other when T_{L1} and T_{RM} are used. This effect does not occur for T_{LMS}. In 3.3 as well as here we also simulated a situation with positive and negative outliers. All methods showed a much smaller bias and MSE then. While the MSE of T_{LMS} is the largest one in a steady state for up to five outliers, in case of a level shift we get almost the same results as for only positive outliers.

3.5 Computation times

The computation time needed is important if many variables are to be monitored simultaneously or other algorithms run concurrently. Table 3 shows the mean times of applying the functionals to 1000 samples of sizes 21 and 31 using a self-written FORTRAN program on a Sun workstation ultra spark with 170 MHz and 320 MB Ram. We remark that for T_{L1} the time depends on the data as the number of iterations needed may vary. In case of a steep trend and an additional level shift the computation time increased to 3.5 (n=21) and 6.2 (n=31) seconds. The time

Figure 2: Steady state: Simulated squared bias (top), variance (middle) and MSE (bottom) for the intercept. L_1 regression x, repeated median $\circ$ and LMS $\triangle$.

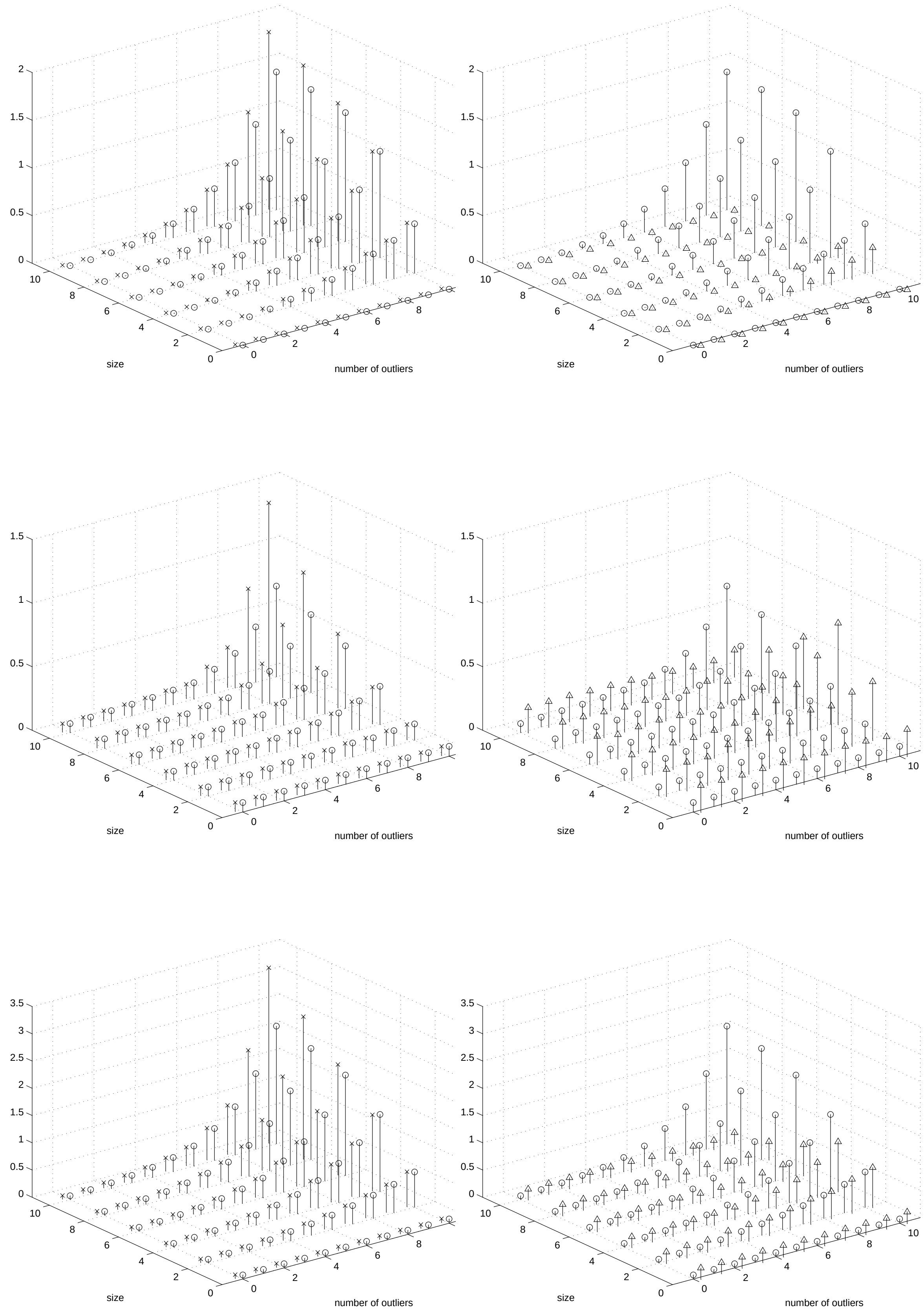
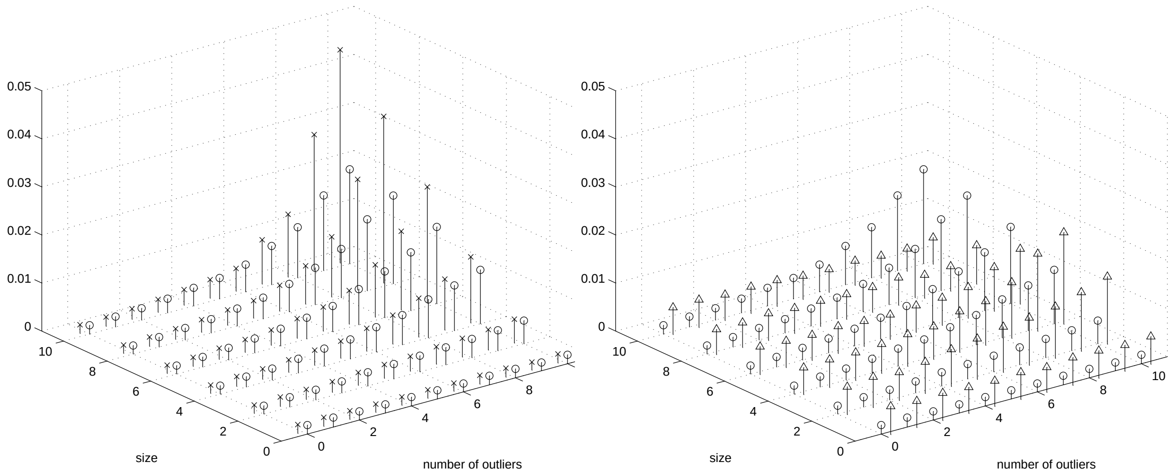


Figure 3: Steady state: Simulated MSE for the slope: L_1 regression x, repeated median $\circ$ and LMS $\triangle$.



needed for computation of T_{LMS} is much larger than that for the other functionals and increases rapidly with the sample size.

Table 3: Mean time of applying the functionals to 1000 samples of different sizes (in seconds) for the steady state $\beta = \mu = 0$.

	T_{L2}	T_{L1}	T_{RM}	T_{LMS}
$n = 21$	0.2	2.4	2.6	28.4
$n = 31$	0.3	4.7	4.4	120.7

3.6 Simulated time series

We now consider a simulated time series of length 250 which is shown in the upper panel of Figure 6. The signal is overlaid with unit noise and 10% of the observations are outliers of size 5. These consist of seven single outliers, four patches of two outliers, two patches of three outliers and one patch of four outliers. The outliers were put at random time points with the exception of the two outliers at time $t = 195$ and $t = 196$ which were put there to make the detection of the level shift at time $t = 201$ more difficult. In order to denoise this time series, we apply T_{L1}, T_{RM} and T_{LMS} using a window width of $n = 31$. We consider edge effects by extrapolating the trend estimated in the first and last time window respectively.

In general, T_{LMS} shows more variability, but it is effected only by the very long outlying pattern at $t = 112$. The most important difference between the methods in

Figure 4: Level shift of size 10: Simulated squared bias (top), variance (middle) and MSE (bottom) for the level. L_1 regression x, repeated median $\circ$ and LMS $\triangle$.

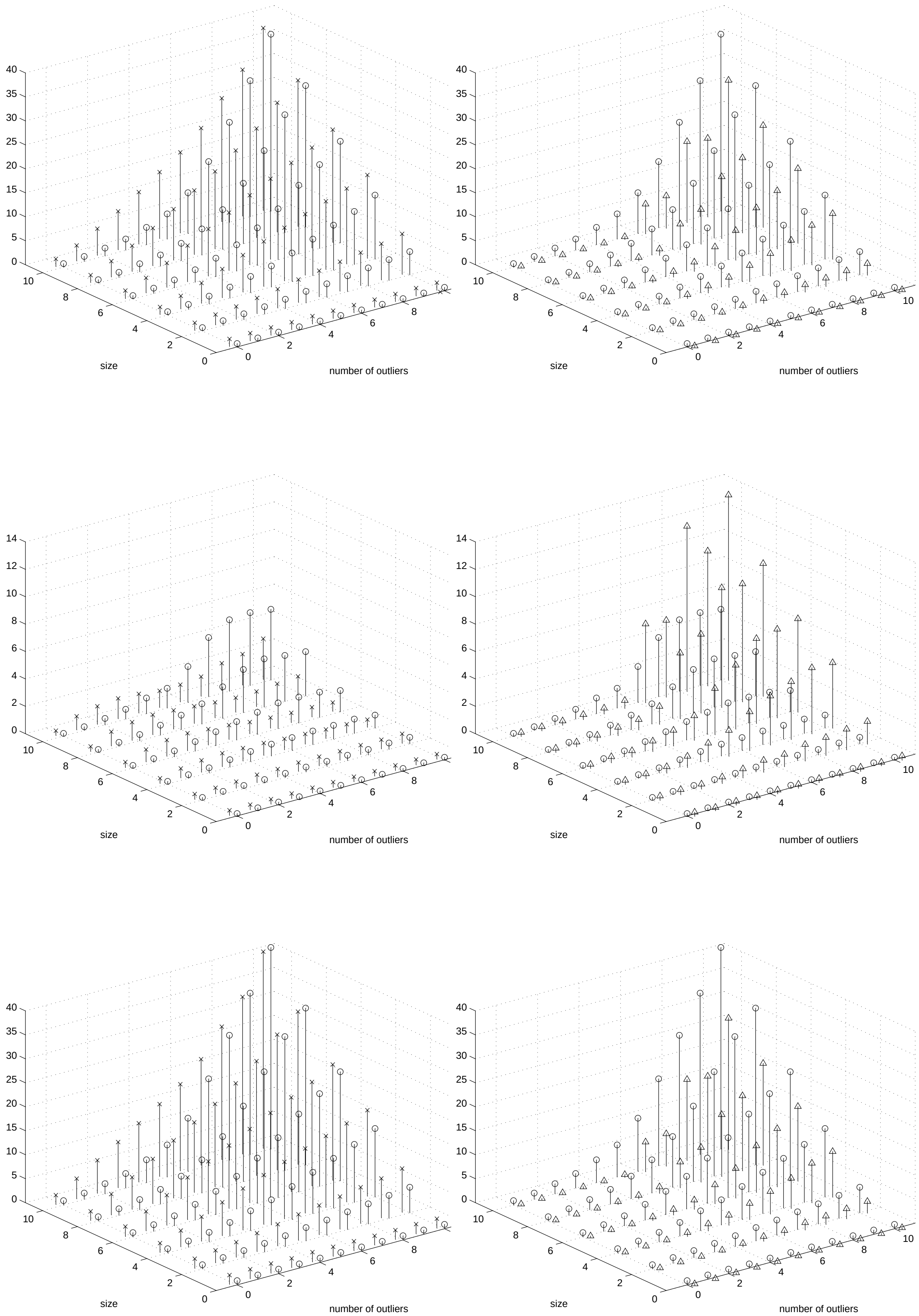
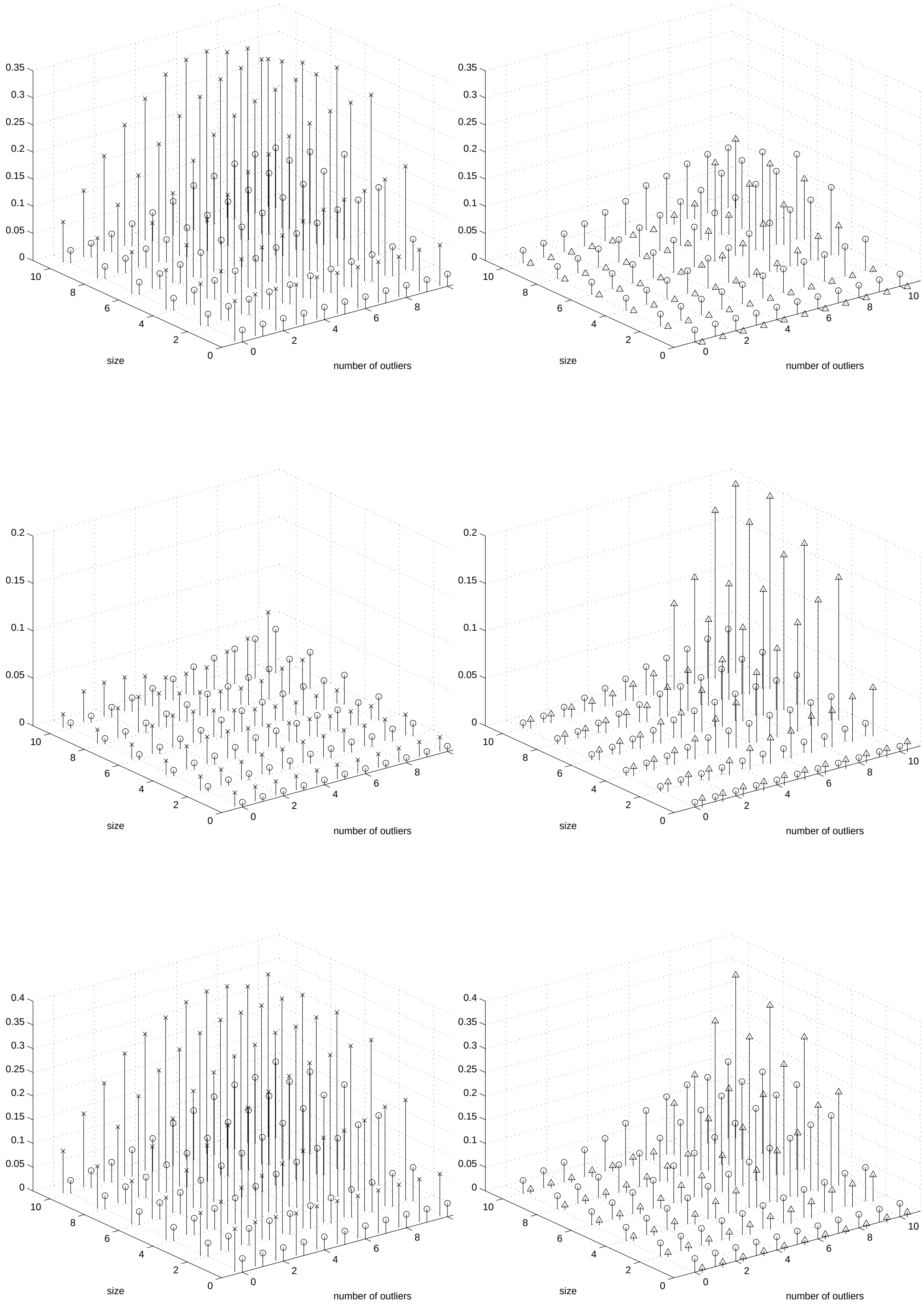


Figure 5: Level shift of size 10. Simulated squared bias (top), variance (middle) and MSE (bottom) for the slope. L_1 regression x, repeated median $\circ$ and LMS $\triangle$.



the clinical context can be seen at the time points of the level shifts. Both T_{L1} and T_{RM} are effected by the level shifts much earlier than T_{LMS} which is first effected only at times $t = 46$ and $t = 194$ respectively. At these time points there are respectively ten and eight observations in the current time window which are effected by the level shift. With respect to the slope the differences between the methods are not very pronounced, but T_{LMS} preserves the slope changes better than the other functionals which tend to smooth the changes.

3.7 Two real examples

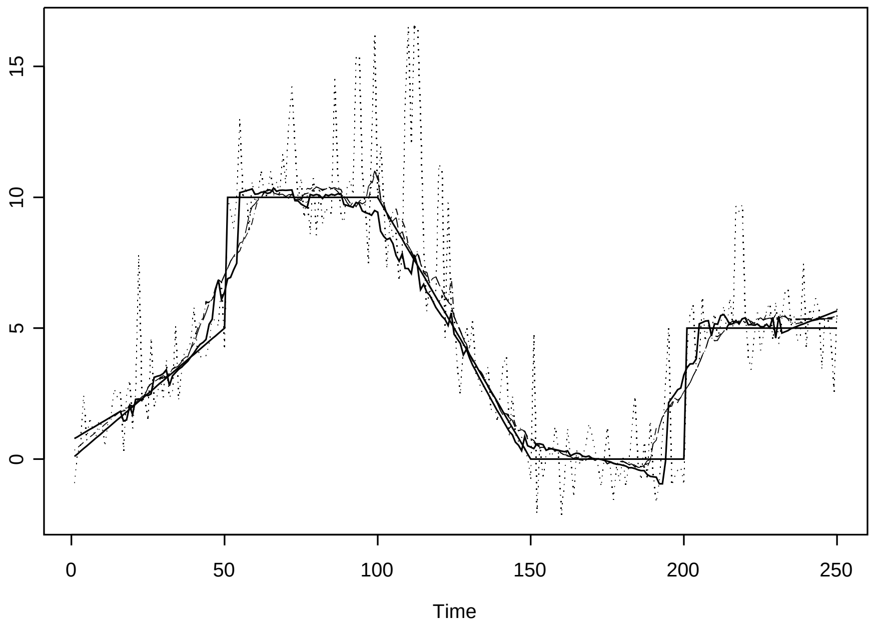
Finally we consider two real examples from the monitoring of intensive care patients. As such data often contain clinically irrelevant minor trends we use a time window of length $n = 31$ (see Figure 7). The first example is the one used in the introduction. T_{L1} and T_{RM} are much less volatile than T_{LMS} which also exhibits a large spike at $t = 63$ due to a particular pattern of outliers. T_{L1} and T_{RM} perform well but overestimate the signal between $t = 110$ and $t = 140$.

The second time series represents the arterial blood pressure of another patient. Again there are outliers but only one section from $t = 225$ to $t = 231$ were judged to be clinically relevant. T_{L1} and T_{RM} are both effected by the clinically irrelevant outliers at about $t = 166$ but miss the clinically relevant outliers at $t = 231$. T_{LMS} performs very well on this data set.

4 Discussion

Alarm systems in intensive care must be capable of on-line detection of clinically relevant patterns such as trends and level changes. The first step in the development of such systems is the on-line extraction of the signal which is corrupted by noise and extreme outliers. In this paper we have compared three robust methods of signal extraction namely T_{L1}, T_{RM} and T_{LMS}. The comparison used simulated and real data as they occur in the monitoring of intensive care patients. On the basis of the limited evidence presented here our tentative conclusions are that T_{LMS} is vary variable and computationally expensive. T_{L1} and T_{RM} can withstand a large number of outliers and are computationally much less expensive. Because of its higher breakdown point the present paper points to T_{RM} as being a prominent candidate for a first sweep over the data in practice.

Figure 6: Simulated time series. Top: Time series (dotted), underlying level (fat solid) and level estimates: T_{L1} (dashed-dotted), T_{RM} (dashed), T_{LMS} (solid). Bottom: Slope estimates (same styles). With respect to the level, T_{L1} and T_{RM} are almost identical.



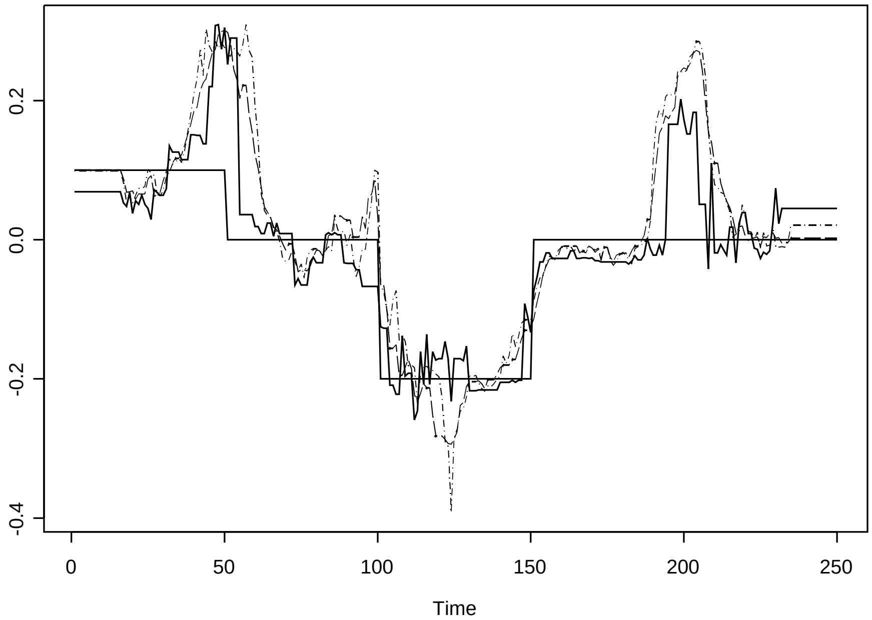
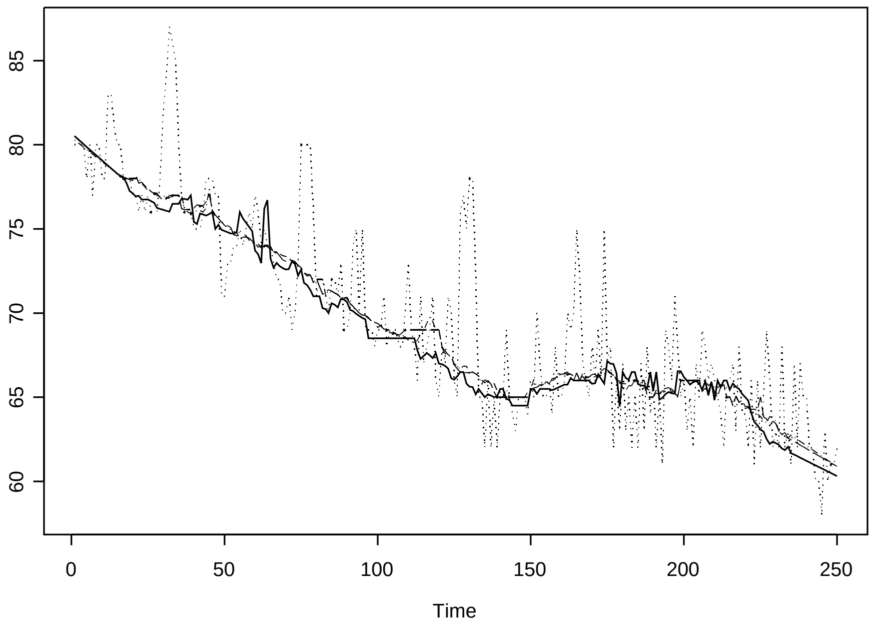
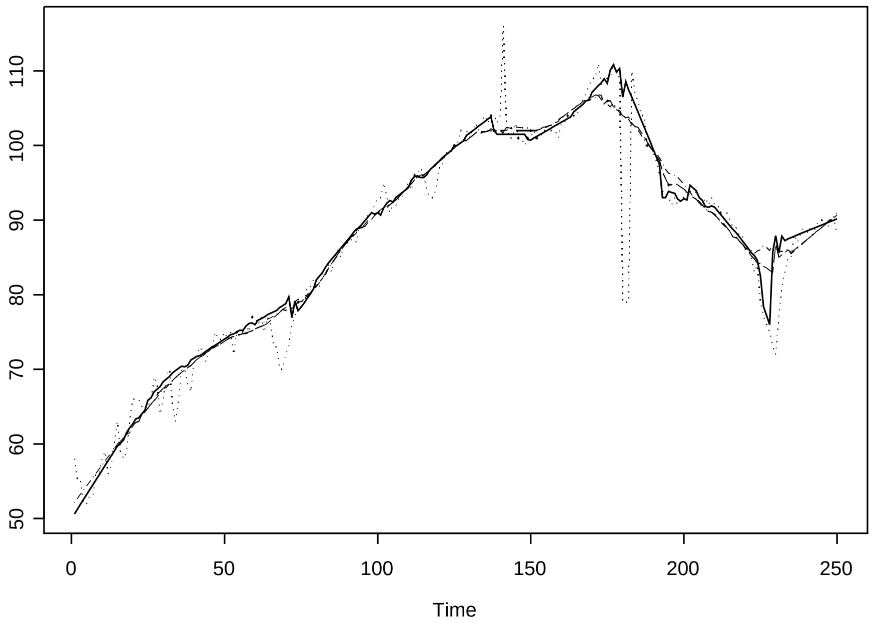


Figure 7: Time series (dotted) representing heart rate (top) and arterial blood pressure (bottom) as well as some level approximates: T_{L1} (dashed-dotted), T_{RM} (dashed), T_{LMS} (solid). T_{L1} and T_{RM} are almost identical.





Acknowledgements:

The financial support of the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of complexity in multivariate data structures") is gratefully acknowledged.

References

Davies, P.L. and U. Gather (1993). The identification of multiple outliers. *J. Amer. Statist. Assoc.* 88, 782–793.

Gather, U., R. Fried and M. Imhoff (2000). Online classification of states in intensive care. In: W. Gaul, O. Opitz and M. Schader, Eds., *Data Analysis. Scientific Modeling and Practical Application.* Springer, Berlin, 413–428.

Gather, U., R. Fried, V. Lanius and M. Imhoff (2002). Online-monitoring of high-dimensional physiological time series - A case-study. *Estadistica*, in press.

Hampel, F.R. (1975). Beyond location parameters: Robust concepts and methods. *Bulletin of the Int. Statist. Inst.* 46, 375–382.

He, X., J. Jureckova, R. Koenker and S. Portnoy (1990). Tail behavior of regression estimators and their breakdown points. *Econometrica* 58, 1195–1214.

Huber, P.J. (1995). Robustness: Where are we now? *Student* 1, 75–86.

Imhoff, M., M. Bauer, U. Gather and D. Löhlein (1998). Statistical pattern detection in univariate time series of intensive care on-line monitoring data. *Intensive Care Medicine* 24, 1305–1314.

Mizera, I. and Ch. Müller (1999). Breakdown points and variation exponents of robust M-estimators in linear models. *Annals of Statistics* 27, 1164–1177.

Rousseeuw, P.J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.* 79, 871–880.

Siegel, A.F. (1982). Robust regression using repeated medians. *Biometrika* 68, 242–244.

Sposito, V.A. (1990). Some properties of L_p-Estimators. In: K.D. Lawrence and J.L. Arthur, Eds., *Robust Regression. Analysis and Applications.* Marcel Dekker, New York, 23–58.

Stromberg, A.J. (1993). Computing the exact least median of squares estimate and stability diagnostics in multiple linear regression. *SIAM J. Sci. Comput.* 14, 1289–1299.

Terbeck, W. (1966). Interaktionen in der Zwei-Faktoren-Varianzanalyse. Doctoral Thesis, Universität Essen, Essen, Germany.

Tukey, J.W. (1977). *Exploratory data analysis.* Addison-Wesley, Reading, Mass., (preliminary ed. 1971).